线上调查平台数据分析模块设计浅谈

李伟男
（中央广播电视总台，北京 100020）

摘　要：本文主要介绍数据分析模块的设计与实现。通过建设在线调查服务平台，实现调查问卷精细化设计、调查问卷自动化收集、调查数据可视化高效展现、调查数据权威精准发布的在线调查全流程管控。通过收集调查服务平台上的数据，面向不同用户提供问题分析、受访人分析、关联分析、时序分析、指数曲线、地区排名等数据分析应用服务。

关键词：线上调查平台；在线调查服务；数据管理分析；模块设计　　**中图分类号：**TP668　　**文献标识码：**A
文章编号：1671-0134（2022）04-154-04　　DOI：10.19483/j.cnki.11-4653/n.2022.04.046
本文著录格式：李伟男 . 线上调查平台数据分析模块设计浅谈 [J]. 中国传媒科技，2022（04）：144-147.

1. 数据分析模块的背景与意义

问卷调查指的是借助结构化问卷，从一定数量的个体样本或人口单位中回收信息的方法。借助问卷这一工具，调研人员能够对社会活动过程进行系统、具体的测量，并运用统计手段对结果进行量化描述和分析。[1] 传统的问卷调查往往存在着数据采集方式比较单一，调查范围相对有限，数据采集上报及时性不够，数据质量核验工作繁琐容易出错、结果展示不够及时等问题。

然而，随着大数据、互联网等技术的发展，亟需通过改变传统的业务流程，采用智能化的数据入库方式、多种形式的可视化展现方式对分析结果进行展现，并为多类型的用户提供服务。[2] 基于互联网技术的媒体占据重要市场，是立足于历史、聚焦于技术革新的发展成果，这种变化的突出表现在于传媒企业建设全面靠近互联网。[3]

通过对收集的数据进行多方位的分析，一方面可以服务于每年一次的大调查工作；另一方面还可以服务于各节目部门的相关调查需求。建设在线调查服务平台，实现调查问卷精细化设计、调查问卷自动化收集、调查数据可视化高效展现、调查数据权威精准发布的在线调查全流程管控。

数据采集。提供对发布的调研问卷的数据采集管理，包括用户 ID 识别认证、扫描方式多样化、防作弊机制、地理位置采集、终端信息采集等。

问卷管理。对调查问卷进行配置，能够生成不同类型、不同样式的调研问卷；提供对调研问卷列表的基础管理、发布等；对当前执行的调研问卷提供进度统计、问卷分析等。

数据分析。面向不同的用户提供问题分析、受访人分析、关联分析、时序分析、指数曲线、地区排名等数据分析应用服务。

数据发布。提供将分析应用结果实时推动至 Web 端、可视化大屏（PC 端）、微信公众号、其他网络新媒体的服务。

会员管理。针对调查对象进行分析，将注册的用户纳入央视大调查的会员管理体系，提供会员信息的查看、消息发布、社区互动、维护管理等。

数据治理。数据治理包括数据标准体系（指标体系、口径标准、标签分类等）、分析指标体系（受访者对象分析、调查主题分析、检索模式等）和数据库建设（主题数据库和专题）三方面的工作。

运维管理包括服务器管理、应用管理、安全管理等。

2. 数据分析模块的需求

数据分析包括基于建模数据的主题分析、数据建模能力、多维度分析能力、数据可视化能力和数据分析管理能力。[4]

2.1　基于建模数据的主题分析

需要支持问题分析、受访人分析、历史分析（时序）、地区排名、指数曲线等适应大调查特点的数据分析。支持分析指标模型及逻辑的可视化定制。

2.2　数据建模能力

数据建模在本系统中的主要是数据挖掘的方法来实现。数据挖掘是按既定目标，对大量数据进行探索和分析，揭示隐藏的、未知的或验证已知的规律性，并进一步将其模型化。[5] 数据建模需支持 Oracle、MySQL、SQL Server、PostgreSQL 等常见的关系型数据库的数据源。支持 Excel、CSV 文件数据源。支持不同数据源的跨 Schema 表关联查询。支持图形化拖拽方式进行数据关联建模，并支持建模数据实时预览。支持自定义字段，可创建新的数据字段。支持字段名更改，并能够自定义字段排序规则。创建的数据集模型支持导出为 Excel 文件。支持手写 SQL 方式创建数据集。

2.3　多维度分析能力

OLAP 技术中比较典型的应用是对多维数据的切片和切块、钻取、旋转等，它便于使用者从不同角度提取有关数据。[6] 通过拖拽形式创建维度层次，要能够调整维度层次中不同层次顺序及命名。支持求和、最大值、最小值、平均值、计数、不重复计数等聚合方式，并支持同环比计算。同时要支持按照升序、降序、自定义规则进行排序，并支持主次排序维度。支持时间类型字段的自动拆分，要能够拆分为年、季度、月、周、日不同粒度。支持过

滤筛选，时间类型字段能够设置时间筛选粒度及时间范围，并且字符类型字段要具备通配符匹配功能。支持联动钻取分析，可以根据设置的维度层次进行上卷、下钻、查看明细功能。

2.4 数据可视化能力

数据可视化技术可以快速的表达繁多的数据和信息进行图像化，从而降低数据的认知难度，帮助人们理解数据。[7] 通过对数据分析挖掘形成的趋势规律，是数据价值的体现。通过数据可视化手段将这些价值呈现出来也是数据实践的重要一环。[8] 大屏所有展示内容都通过工具配置，不需要编写代码，能够做到快速构建，快速变更，维护方便。支持设置分辨率大小或自适应，固定分辨率下，支持按照原始大小或缩放长边铺满展现。支持故事板播放运行，需支持两种播放效果，一是一屏展现所有内容，二是多屏翻页展现。

2.5 数据分析管理能力

数据管理模块支持 URL 分享，并支持公开、密码私密分享两种方式。支持对组织结构、用户、角色进行统一的管理，支持与其他系统进行对接实现单点登录等功能。

3. 数据分析模块的设计方案

3.1 基于建模数据的主题分析

如图 1 所示，建立基于问题分析、受访人分析、历史分析（时序）、地区排名、指数曲线等适应大调查特点的数据分析接口。

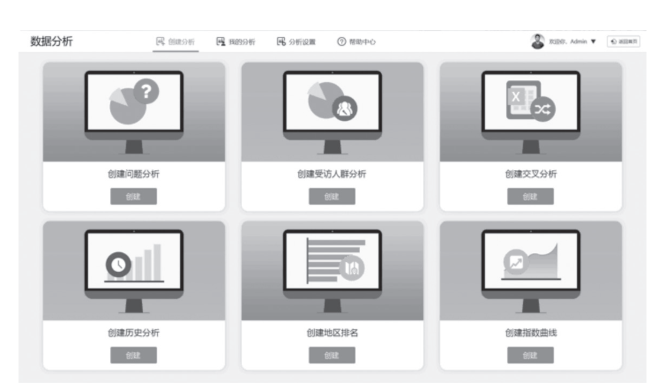


图 1 基于主题的数据分析界面

此外，还设计了支持分析指标模型及逻辑的可视化定制页面，如图 2 所示。

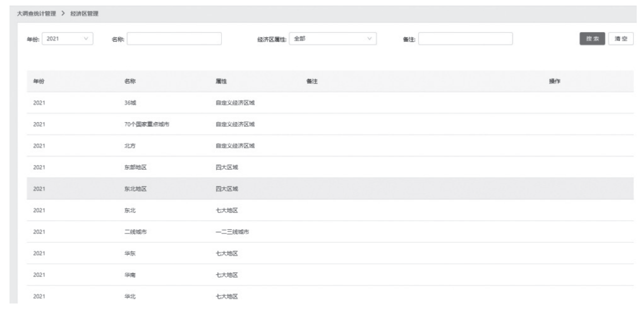


图 2 数据分析列表

3.2 数据模型的设计

针对数据分析模型需要支持多种不同的数据源，设计如图 3 所示的多数据源类型的数据接入或存储方案。该设计方案支持可视化快速创建数据集，用户可拖拽数据源中多个数据表搭建数据关联关系，数据表间通过连接线清晰展现数据关联关系。如果两个表之间存在主外键关联，会自动建立关联关系，如果没有会根据相同的字段名称建立关联，支持关联关系的修改。支持自定义字段功能，支持以下五大类：时间计算、数字计算、字符计算、关系计算和逻辑计算。支持通过数据可视化窗口实时查看勾选的字段以及自定义字段的数据内容及格式。

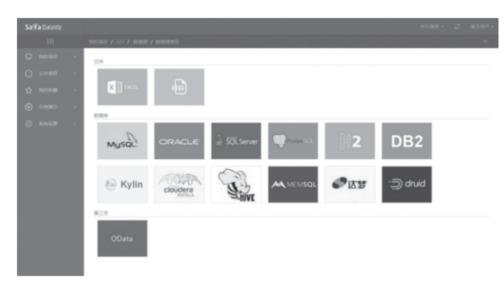


图 3 数据源类型

针对高级需求还支持可以手写 SQL 方式，实现数据建模过程，如图 4 所示。创建的数据集模型支持导出为 Excel 文件。支持手写 SQL 方式创建数据集。



图 4 数据建模

3.3 数据多维度分析设计

如图 5 所示，设计支持多种数据聚合方式并支持同环比计算的数据多维度分析模块。用户可以随时获得数据不同维度的分析结果。

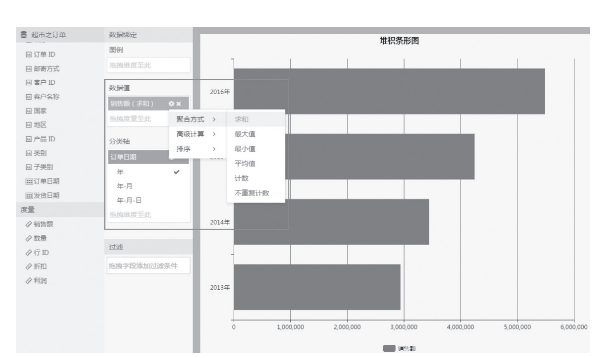


图 5 数据聚合方式设计

此外，支持多种排序方式，排序方式主要包括排序顺序、Top 数据、排序依据、排序提示几个模块，并支持多字段排序叠加。

升序：升序为当前排序的正常升序排序。当对度量字段进行排序时，为当前度量的数字升序；当对维度字段进行排序时，为当前维度排序依据的升序。

降序：降序为当前排序的正常降序排序。当对度量字段进行排序时，为当前度量的数字降序；当对维度字段进行排序时，为当前维度排序依据的降序。

Top 数据：Top 数据功能只对度量有效，支持两种选择方式，当选择排序顺序为升序时，UI 变为“前 __ 个”，当选择排序顺序为降序时，UI 变为“后 __ 个”。

默认排序：即为数据查询的原始顺序，配合排序顺序选择是升序（正序）或降序（逆序）。

字母 / 拼音顺序：按照当前字段的字母 / 拼音顺序进行排序。

其他字段排序：选择其他的字段来排序选择的字段，并支持聚合，如图 6 所示，当选择字段后，需要同时选择该字段的聚合方式，并配合排序顺序，以聚合后的结果作为依据来排选择的字段。

手动排序：手动排序即支持用户手动调整排序规则，系统默认读取选择排序字段的 distinct 值，用户可以自行拖拽、添加、删除字段来调整先后顺序，然后形成排序规则，并按照此规则排图表数据。

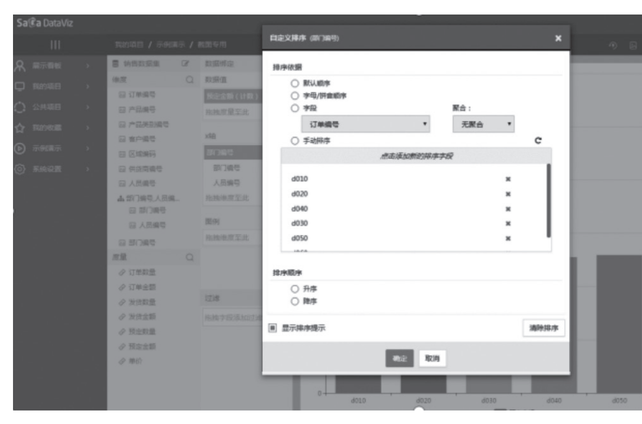


图 6 字符类型过滤条件

该模块提供多维度探索分析功能，用户可根据数据分析需求自由拖拽数据集中的维度和度量完成数据绑定，并自动生成对应可视化图表。在数据分析阶段还支持对字段的二次加工，公式种类分为六大类：时间计算、数字计算、字符计算、关系计算、逻辑计算、聚合计算。对于分组数据，还支持拖拽维度改变分组层次，并进行组合分析。支持数据上卷、下钻、钻透、切片、切块，可以完成从粗粒度数据到明细数据的逐层探查，并自动生成对应可视化图表，帮助用户发现业务根本问题，如图 7 所示。



图 7 数据钻取

3.4 数据可视化设计

可视化展示离不开图表，提供数 10 种具有动态交互性可拓展的图表，如图 8 所示。

支持基于 GIS 地图的数据可视化方法，包括轨迹图、热力图、标记点地图，GIS 地图支持在线的高德地图、必应地图、谷歌地图等，也可支持基于 WMS、WMTS 标准的非在线地图产品，GIS 地图底图可随意切换。

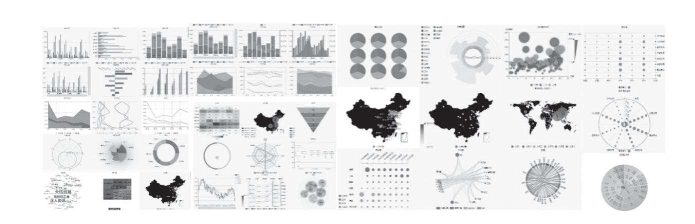


图 8 可视化图表

支持网格式布局，将上述元素拖拽到上、下、左、右、中等位置，并支持通过拖拽的方式改变某一区域的大小；支持使用自由块完成图册的无网格约束的自由布局。支持设置图表间联动交互，并且支持跨数据源联动，帮助用户动态分析数据、深入探索数据间复杂关系，如图 9 所示。

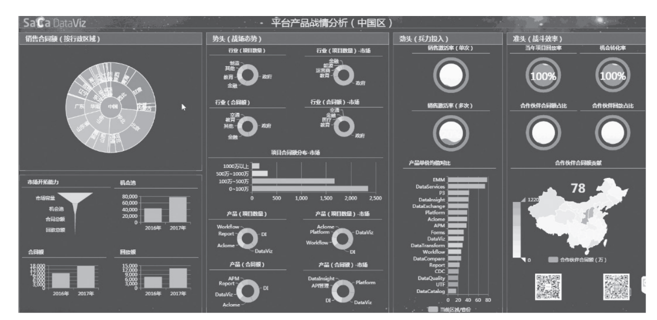


图 9 联动展示

3.5 数据分析管理设计

内置多种主题风格样式，支持一键快速切换各种主题，针对于图册主题有更为特殊的要求，还支持主题的自定义功能，通过图形界面的方式增加自定义主题。针对于一个图册中的多个图表分别使用不同主题的需求，

提供在图册中图表保留本身主题的功能，即图表可以不使用图册的主题，如图 10 所示。



图 10 图册主题

4. 数据分析模块的应用实践

通过建设在线调查平台的数据多维度分析模块，可以提供数据分析结果应用，以及分析结果并未公布。一方面可以适应多种不同数据源，建立相应的数据集，另一方面，可以提供丰富的可视化界面，用户可以从不同的维度了解数据分析的结果。

5. 下一步工作

在未来的工作中，基于在线调查服务平台，扩展数据分析结果的应用，融合最新的指数分析指标算法。同时，可以接入移动端、微信微博等社交媒体，将数据分析结果直接发布。媒

参考文献

[1] 邵国松，谢 珺 . 我国网络问卷调查发展现状与问题 [J]. 湖南大学学报（社会科学版），2021（4）：149–155.

[2] 施俊 . 数据仓库技术在电视台节目播出质量控制中的应用分析 [J]. 数字通信世界，2019（11）： 68.

[3] 颜 铮 . 媒体融合背景下传媒业态变化研究 [J]. 中国传媒科技，2021（7）： 79–81.

[4] 陈冬 . 徐州广播电视台新媒体营销策略研究 [D]. 徐州：中国矿业大学，2019.

[5] 古欣 . 大数据分析数据建模的思路 [J]. 统计与咨询 ,2017(03):57–59.

[6] 徐涛 , 段雪莹 . 多维数据综合分析评价方法研究 [J]. 科技信息 (科学教研),2008(05):73–75.

[7] 郭思思 . 以用户为中心的社交网络数据可视化设计方法研究 [D]. 湖南大学 ,2018.

[8] 袁然 . 全媒体传播中数据技术的应用实践 [J]. 中国传媒科技，2021（7）： 21–23.

作者简介：李伟男（1988–），男，北京，本科，工程师，中央广播电视总台，研究方向：大数据数据分析。

（责任编辑：张晓婧）

（上接第143页）

一个重要议题。

最后，虽然智能化、自动化可在一定程度上提高新闻制作效率，但自动化新闻系统永远无法替代编辑、记者的工作。再多的自动化系统也无法取代和构建讲述好故事的技能。在这里，自动化系统为讲好故事提供了必要的支持，而不是取代它。明确了这两者的关系定位，也就更加清楚地认识到在构建自动化系统的实践中，仅有工程师的参与是不够的。正如英国卫报数据博客编辑西蒙・罗杰斯所指出的：新闻编辑部的布局很有讲究，如果你越靠近新闻编辑部，就更加方便进行新闻报道的交流，这是新闻策划过程中必不可少的一部分；反之，则两者距离越来越远。这也从另一方面阐明：专家型新闻人才与技术团队相互协作，是新闻自动化系统成功构建的必要条件。比如，在上述开发自动化图表新闻项目中，项目组还邀请了认知科学家、漫画理论家、一线记者共同参与项目中，提出各自专业的见解。

综上所述，利用人工智能技术，特别是自然语言处理技术提升新闻报道时效性是时代的必然。新闻媒体要适应新时代新闻制作、传播的特点，利用新技术切实履行好信息传播、舆论引导的使命。同时，新闻人和技术工程师的密切合作是开启通往新闻智能化大门的金钥匙。媒

参考文献

[1] Daniel Faggella. Artificial Intelligence for Content Marketing and Content Creation.https：//podcast.emerj.com/artificial–intelligence–for–content–marketing–and–content–creation.2018.2.16：1–2

[2] Samuel Danzon–Chambaud.Covering COVID–19 with automated news. The Tow Center for Digital Journalism.2021.8.6：2

[3] Nic Newman.digital news report 2021 10th edition.Reuters Institute.2021：1

[4] Jan Bruck.How BBC World Service fights the coronavirus infodemic with chat bots. https：//bbcnewslabs.co.uk/news/2021/ws–corona–bot/.2021：2

作者简介：张建伟（1980–），男，北京，新华社北美总分社高级工程师，研究方向：新闻自动化。

（责任编辑：李净）